Appendix A: Comparison of Proposed Measures

Measurement: **Percent Blockage on Interconnection Trunks**

Reference: **VI. A. Trunk Blockage Measurements**

	[illegible]	[illegible]	[illegible]
Measurement Name	Percent Blockage on Interconnection Trunks	Ameritech is proposing alternative measures: ▪ Call Completion Rates However, Ameritech will provide the following measures if trunk blockage must be reported: ▪ Trunking Grade of Service (Final Trunk Groups)	Ameritech proposes call completion measures rather than percent blockage.
Calculation	Final Interconnection Trunk Groups Blocked During Reporting Period/Total Number of Interconnection Trunk Groups	Call Completion: [(Number of Call Attempts – Number of Blocked Calls + Number of Successful Reroutes) / Total Call Attempts] * 100 Ameritech will provide the following calculation if trunk blockage must be reported: ▪ Percentage of EOI Final Trunk Groups Carrying Local and Interlata/Intralata Toll Traffic from an Ameritech Tandem to a CLEC End Office that Experience Blockage Above a Specified Threshold during the Average Busy Hour ▪ Intralata: 1% Threshold ▪ Interlata: 0.5% Threshold	Ameritech proposes a call completion calculation rather than percent blockage.
Exclusions and/or Inclusions	▪ None	Call completion proposed exclusions: ▪ Blockage that results from actions or failures to act on the part of the CLEC ▪ New CLECs from overall measures ,for all CLECs during an initial six-month period when they are being established (Note: These are reported separately) Call completion proposed inclusions: ▪ Final trunk groups that are not designed to overflow to other trunk	▪ Ameritech proposes additional exclusions and inclusions as stated in the previous column. ▪ New CLECs are not included in overall measures during an initial six-month period, so that the measure is not biased by results of carriers that have not yet established their networks and ascertained their needs.

Appendix A: Comparison of Proposed Measures

		groups when trunk blockage occurs ▪ Traffic that is actually rerouted and completed ▪ Measurement on a 24-hour day basis for a specified number of days per month, such as twenty (20) business days	
Categories (Wholesale)	▪ Interconnection Trunks ▪ Common Trunks	▪ Interlata ▪ Intralata	
Categories (Retail)	▪ Common Trunks	▪ Interlata ▪ Intralata	

Appendix A: Comparison of Proposed Measures

Measurement: **Percent Blockage on Common Trunks**

Reference: **VI. A. Trunk Blockage Measurements**

	[illegible]		
Measurement Name	Percent Blockage on Common Trunks	Ameritech is proposing alternative measures: ▪ Call Completion Rates However, Ameritech will provide the following measures if trunk blockage must be reported: ▪ Trunking Grade of Service (Final Trunk Groups)	Ameritech proposes call completion measures rather than percent blockage.
Calculation	Final Common Trunk Groups Blocked During Reporting Period/Total Number of Common Trunk Groups	Call Completion: [(Number of Call Attempts – Number of Blocked Calls + Number of Successful Reroutes) / Total Call Attempts] * 100 Ameritech will provide the following calculation if trunk blockage must be reported: ▪ Interlata: Percentage of Alternate Final Trunk Groups Which Are Block at a Rate of 0.5% or More during the Average Busy Hour ▪ Intralata: Percentage of Direct Final Trunk Groups Which are Blocking at a Rate of 1% or More during the Average Busy Hour	Ameritech proposes a call completion calculation rather than percent blockage.
Exclusions and/or Inclusions	▪ None	Call completion proposed exclusions: ▪ Blockage that results from actions or failures to act on the part of the CLEC ▪ New CLECs from overall measures ,for all CLECs during an initial six-month period when they are being established (Note: These are reported separately) Call completion proposed inclusions: ▪ Final trunk groups that are not designed to overflow to other trunk groups when trunk blockage occurs	▪ Ameritech proposes additional exclusions and inclusions as stated in the previous column. ▪ New CLECs are not included in overall measures during an initial six-month period, so that the measure is not biased by results of carriers that have not yet established their networks and ascertained their needs.

Appendix A: Comparison of Proposed Measures

		▪ Traffic that is actually rerouted and completed ▪ Measurement on a 24-hour day basis for a specified number of days per month, such as twenty (20) business days	
Categories (Wholesale)	▪ Interconnection Trunks ▪ Common Trunks	▪ Interlata ▪ Intralata	
Categories (Retail)	▪ Common Trunks	▪ Interlata ▪ Intralata	

Appendix A: Comparison of Proposed Measures

Measurement: **Average Time to Respond to a Collocation Request**

Reference: **VI. B. Collocation Measurements**

	Commission	Ameritech	VARIATIONS
Measurement Name	Average Time to Respond to a Collocation Request	Ameritech proposes a similar measure to the NPRM measure, but with the modifications indicated below.	Ameritech agrees to add this measure.
Calculation	[Σ[(Request Response Date and Time) - (Request Submission Date and Time)]]/Count of Requests Submitted in Reporting Period	Ameritech proposes a similar calculation to the NPRM measure, but with the following modification: ▪ The date of submission is the date that the request is received and the date of completion is the date that the response is sent out by the incumbent LEC. ▪ The measure is reported using the date only, not the time.	The NPRM proposes that the measure be calculated using date and time, while Ameritech proposes the measure using days only.
Exclusions and/or Inclusions	Exclusions: ▪ Orders cancelled by competing carrier	Proposed exclusions: ▪ Orders cancelled by competing carrier are excluded from the calculation Proposed inclusions: ▪ The clock for each measure is restarted if the CLEC modifies its request. ▪ The clock stops when the incumbent LEC sends out the CLEC a response providing space availability and cost information.	▪ Ameritech proposes additional exclusions and inclusions as stated in the previous column.
Categories (Wholesale)	▪ Physical collocation ▪ Virtual collocation	▪ Ameritech does not believe that this measure applies to virtual collocation	▪ Ameritech does not believe that this measure applies to virtual collocation
Categories (Retail)	▪ No Equivalent	▪ No equivalent	▪ No variation

Appendix A: Comparison of Proposed Measures

Measurement: **Average Time to Provide a Collocation Arrangement**

Reference: **VI. B. Collocation Measurements**

	Commission	Ameritech	VARIATIONS
Measurement Name	Average Time to Provide a Collocation Arrangement	Ameritech proposes a similar measure to the NPRM measure, but with the modifications indicated below.	Ameritech agrees to add this measure.
Calculation	[∑[(Date and Time Collocation Arrangement is Complete) - (Date and Time Order for Collocation Arrangement Submitted)]]/Total Number of Collocation Arrangements Completed During the Reporting Period	Ameritech proposes a similar calculation to the NPRM measure, but with the following modification: ▪ The date of submission is the date that the request is received and the date of completion is the date that the response is sent out by the incumbent LEC. ▪ The measure is reported using the date only, not the time.	The NPRM proposes that the measure be calculated using date and time, while Ameritech proposes the measure using days only.
Exclusions and/or Inclusions	Exclusions: ▪ Orders cancelled by competing carrier	Proposed exclusions: ▪ Orders cancelled by the competing carrier are excluded from the calculation. ▪ The incumbent LEC is not held accountable for any CLEC delays in arranging final walkthrough or accepting the space. ▪ Requests that relate to interconnection agreements with specified due dates are excluded from the calculation. Proposed inclusions: ▪ The clock for each measure is restarted if the CLEC modifies its request.	▪ Ameritech proposes additional exclusions as specified in the previous column.
Categories (Wholesale)	▪ Physical collocation ▪ Virtual collocation	▪ Ameritech concurs with categorization	▪ No variation
Categories (Retail)	▪ No Equivalent	▪ No retail equivalent	▪ No variation

Appendix A: Comparison of Proposed Measures

Measurement: **Percent of Due Dates Missed with Respect to the Provision of Collocation Arrangements**

Reference: **VI. B. Collocation Measurements**

	Commission	Ameritech	Variations
Measurement Name	Percent of Due Dates Missed With Respect to the Provision of Collocation Arrangements	Ameritech proposes a similar measure to the NPRM measure, but with the modifications indicated below.	Ameritech agrees to add this measure.
Calculation	[Number of Orders Not Completed within ILEC Committed Due Date During Reporting Period/Total Number of Orders Scheduled for Completion in Reporting Period] x 100	Ameritech proposes a similar calculation to the NPRM measure, but with the following modification: ▪ The date of submission is the date that the request is received and the date of completion is the date that the response is sent out by the incumbent LEC. ▪ The measure is reported using the date only, not the time.	The NPRM proposes that the measure be calculated using date and time, while Ameritech proposes the measure using days only.
Exclusions and/or Inclusions	Exclusions: ▪ Orders cancelled by competing carrier	Proposed exclusions: ▪ Orders cancelled by competing carrier are excluded from the calculation. Proposed inclusions: ▪ The clock for each measure is restarted if the CLEC modified its request.	▪ Ameritech provides additional exclusions and inclusions as stated in the previous column.
Categories (Wholesale)	▪ Physical collocation ▪ Virtual collocation	▪ Ameritech concurs with categorization	▪ No variation
Categories (Retail)	▪ No equivalent	▪ No retail equivalent	▪ No variation

Before the
FEDERAL COMMUNICATIONS COMMISSION
Washington, D.C. 20554

In the Matter of:	)	
	)	
Petition of LCI and CompTel for	)	CC Docket No. 98-56
Expedited Rulemaking To Establish	)	RM-9101
Reporting Requirements and	)	
Performance and Technical Standards	)	
for Operations Support Systems	)	

APPENDIX B TO COMMENTS OF AMERITECH

John T. Lenahan
Larry A. Peck
AMERITECH CORPORATION
30 South Wacker Drive
Chicago, Illinois 60606
(312) 750-5000

Theodore A. Livingston
John E. Muench
Christian F. Binnig
Demetrios G. Metropoulos
MAYER, BROWN & PLATT
190 South LaSalle Street
Chicago, IL 60603
(312) 782-0600

COUNSEL FOR AMERITECH

June 1, 1998

TABLE OF CONTENTS

PAGE

Before the
FEDERAL COMMUNICATIONS COMMISSION
Washington, D.C. 20554

In the Matter of:	)	
	)	
Petition of LCI and CompTel for	)	CC Docket No. 98-56
Expedited Rulemaking To Establish	)	RM-9101
Reporting Requirements and	)	
Performance and Technical Standards	)	
for Operations Support Systems	)	

I. Qualifications

1. The author of this paper is Daniel S. Levy, who has a Ph.d. in Economics from The University of Chicago, and who serves as the Director of the Economics Practice for the Central Region consulting office of Arthur Andersen LLP. The Central Region consulting offices includes Chicago, Cincinnati, Cleveland, Detroit, Minneapolis, and St. Louis, among other locations.

II. Introduction

2. The purpose of this paper is to respond to the Commission's request for comments in its Notice of Proposed Rulemaking on performance measurements (the "Notice"). Specifically, this paper is directed to the statistical analyses proposed by the Commission for the analysis of performance measurement data contained in Appendix B of the Notice.

3. This paper discuses several statistical tests that can be used to determine whether incumbent LEC's are providing substantially similar service to both customers of CLECs and incumbent LECs. The general framework proposed is consistent with one presented by AT&T in their *ex parte* communication to the FCC dated February 3, 1998. In that document AT&T points out that the performance of the incumbent LEC on any given observed performance measure represents a specific outcome of a process that contains a random component. The observed performance of an incumbent LEC on any given measure will change from one period to the next even if the underlying performance of the incumbent LEC is consistent. Similarly, even though the incumbent may be providing equal levels of service to both its own and CLEC customers, random variation and chance will result in differences in the measured service received by CLEC and incumbent customers during any given measurement period. With the essential changes discussed below, I believe that an overall parity test described by AT&T, based on the number of individual performance measures that are observed to be disparate in a given period, can be used effectively to test for overall parity of service.

4. The statistical methods discussed in this paper can be used to distinguish between differentials in performance generated by random chance and those generated by possible disparate treatment on the part of the LECs. In addition, the statistical analyses and protocols described in this paper can be used to determine the extent to which differences in the composition of the customer base between CLECs and a LEC is responsible for apparent differences in service provided by the LECs. Where disparity may appear to

exist at a highly aggregated level a more appropriate level of disaggregation may show that parity exists.

5. I therefore suggest that a multiple stage protocol be used to check for parity. In the first stage a pre-specified set of standard statistical techniques should be used to assess parity. If this first stage analysis demonstrates parity, no further analysis will be required. The statistical tests and level of disaggregation of the performance measures used in this first stage will need to be determined prior to testing. This first stage of statistical testing is described below.

6. Because of the complexity of the factors that affect the services that the incumbent LECs provide, it is likely that on occasion these standard tests will indicate a lack of parity when in fact parity does exist. I suggest when on indication of possible disparity is observed in the first stage analysis, a second stage of analysis should be performed to determine the source of the apparent disparity. In some cases, the apparent disparity will be attributable to some factor that does not reflect disparate service, but rather results from some acceptable market or service factor that was not reflected in the first stage analysis. In other cases, disparity may exist and be more specifically identified by this analysis. This second stage analysis may use a broad range of sophisticated statistical techniques to help pinpoint the cause of disparity.

7. The statistical analyses and testing protocols that are outlined in this paper are based on the assumption that if parity is not observed, the first course of action should be to investigate and correct the problem. If fines or other punitive actions are the immediate result of a mere indication of possible disparity, alternative testing protocols and

investigation would need to be developed. The statistical analysis described in this paper are designed with the intent that if disparity is observed, the LEC and the CLEC will work quickly to identify and correct the source of the disparity. The testing protocol described below would not be appropriate in the setting were significant fines or other punitive actions were taken against an incumbent LEC that was moving quickly to correct apparent or real disparities in service.

III. Statistical Methods

8. Variables that may provide evidence of discrimination should be divided into two categories: continuous variables (i.e., variables which can take on any value) and Boolean variables (variables which can take on either of exactly two values, typically 0 and 1). Variables which can take on only certain positive integer values (e.g., 0, 1, or 2) may also be used. The calculations below may be somewhat more complicated in such cases, unless one can either re-state the variables as Boolean variables or unless the variables can take on enough values to be approximately continuous.

9. There are two significant steps to a proper analysis of performance measures. The first is to examine differences between comparable performance measurements for incumbent and competing local exchange carriers, in order to identify possible instances of non-parity or discrimination. Once these areas of concern are identified, the second step is to investigate and analyze the differences in greater detail so as to determine whether discrimination has occurred, or whether the observed differences in performance data arise from another source.

A. Step 1: Identification of Possible Discrimination

10. For continuous variables, the following measures can be calculated:

(1) Z-Statistic for Difference in Means = ratio of difference in means to the standard deviation of the difference. A value that exceeds the 95 percent confidence level should be viewed as an indication of possible disparity.[1/]

(2) Ratio of Variances: The ratio of the sample variances will be distributed with an F-distribution if the performance measure is normally distributed. Values of this ratio outside of the critical value of the appropriate F-distribution should be viewed as an indication of possible disparity.

11. For Boolean variables, only a test for difference in means need be calculated.

12. Suppose there are two variables, x1 and x2, which measure the variable for the incumbent LEC and competing LECs, respectively, with n1 and n2 observations of these two variables, respectively. Call the sample means of these two variables m1 and m2, respectively, and the sample variances of these two variables v1 and v2, respectively.

13. In order to measure the statistical significance of the difference in the means of these two variables (i.e., m2 - m1) it is necessary to make assumptions regarding the variances of the populations from which x1 and x2 are drawn, as well as the shape of the distribution of x1 and x2.

[1/] Statistical tests based on the Z-Statistic are generally not recommend for samples of less than roughly 30 observations. The analysis I propose below is based on sample sizes of at least 30. If tests based on less than 30 observations are used, other tests such as t-statistic should be considered.

14. The simplest case is if one is willing to assume that x1 and x2 are normally distributed with the same underlying population variance. In this case, one can apply a simple parametric test of the equality of the means. The parametric test is described as follows.

1. Parametric Tests

a. Continuous Variables

15. If x1 and x2 are continuous variables (i.e., may take on any value) and are normally distributed, and the population variance (σ^2) is equal for x1 and x2, then the variance of the difference in the means, (m2 - m1), is equal to the following:

$$Var(m2 - m1) = [v2 \bullet (n2 - 1) + v1 \bullet (n1 - 1)] \bullet \{[(1 / n1) + (1 / n2)] / (n1 + n2 - 2)\}$$ [2/]

16. If all of the above assumptions hold, except that the population variances differ, then the variance is still definable, although its calculation is complicated somewhat. Assuming that x1 and x2 are normally distributed, it is possible to test the assumption that the population variance is equal. This is described below.

17. If x1 and x2 are continuous variables but are not normally distributed, then it may not be possible to calculate the variance of the mean of the two variables. The problem of non-normal variables is discussed below with regard to sample size and the Central Limit Theorem.

b. Boolean Variables

18. For variables that take on either of two values, 0 or 1, the variance of the difference, (m2 - m1), is equal to the following:

2/ David K. Hildebrand and Lyman Ott, Statistical Thinking for Managers, (Duxbury Press, Boston) 1987, p.312.

$$Var(m2 - m1) = m(1-m)\cdot[(1 / n1) + (1 / n2)]^{3/}$$

where m equals the total number of ones in the two samples divided by the total number of observations in the two samples (i.e., is equal to the mean for the two samples combined).

19. For either continuous or Boolean variables, the ratio of the difference in the means, (m2 - m1), to the standard error of this difference (equal to the square root of the variance of the difference) will be distributed following a Student's t-distribution with (n1 + n2 - 2) degrees of freedom, in large samples. In large enough samples, the Z-Statistic may be used for test of means.

c. Testing of Assumptions Necessary for Using Parametric Test

20. The parametric test, as described above, relies upon two assumptions:

(i) that the sample statistics of interest, that is, either the sample mean of the continuous variable, or the sample proportion, m, are normally distributed, and

(ii) that the population variances for the populations from which x1 and x2 were sampled are equal

If the latter assumption does not hold, then a parametric test is still possible, although it is considerably more complicated.

3/ Paul Newbold, Statistics for Business and Economics, 4th Edition (Prentice Hall, Englewood Cliffs, NJ), 1994, P.360. There are additional sample size restrictions associated with this test. In some cases, the test may require more than 30 observations in each population.

d. Testing if Variables are Normally Distributed

21. A normal distribution is uniquely defined by its mean and variance. What this means is that the skewness and kurtosis of a normal distribution are equal to zero. One method for testing whether a variable is normally distributed is to calculate the following statistic:

$$L = n \bullet [(m3^2 / 6) + \{(m4-3)^2 / 24\}]$$ [4/]

where n is the number of observations in the sample, m3 is sometimes called the skewness coefficient and is equal to the following:

$$m3 = \Sigma(x-\bar{x})^3 / [n \bullet s^3]$$

where

$$s = [\Sigma(x-\bar{x})^2 / n]^{1/2}$$

and (m4 -3) is called the degree of excess, where m4 is equal to the following:

$$m3 = \Sigma(x-\bar{x})^4 / [n \bullet s^4]$$

The test statistic L follows a chi-squared distribution with two degrees of freedom where sample sizes are large. Critical values required for these tests in smaller samples are available in standard texts.[5/]

22. For large enough samples, the normality assumption will approximately hold, regardless of the true distribution of the underlying population. This is due to the Central Limit Theorem. The Central Limit Theorem states that for a large enough sample, the sample

---

[4/] Paul Newbold, Statistics for Business & Economics, 4th Edition (Prentice-Hall, Englewood Cliffs, NJ), 1994, pp. 412-414.

[5/] Paul Newbold, Statistics for Business & Economics, 4th Edition (Prentice-Hall, Englewood Cliffs, NJ), 1994, p. 412-415.

mean of any variable with finite mean and variance, expressed as deviations from the population mean, is approximately normal. Hence, for large enough samples, it may not even be necessary to test for normality before proceeding with the battery of parametric tests described here.

e. Differences in Variances

Continuous Variables

23. Using the notation from the above example, the ratio of the sampled variances of normally distributed variables, i.e., v_1 / v_2, has an F-distribution with $(n_1\text{-}1, n_2\text{-}1)$ degrees of freedom if the true variances of the populations from which the two samples are drawn are equal, and the variable is normally distributed.

Boolean Variables

24. The variance of a boolean distribution of n observations with a mean of m is equal to the following:

$$m \bullet (1\text{-}m) / n$$

In other words, the variance of a Boolean variable is a direct function of the mean of the variable. Hence, if the means of two Boolean variables are equal, then so are their variances. In other words, there would be no purpose in testing the difference in the variances of two Boolean variables.

f. Appropriateness of Testing Equality of Variances

25. The Commission seeks comment on whether a test of the equality of variances might be useful as a measure of discrimination. Broadly interpreted, the Commission seeks to

answer whether the distribution of a particular variable is the same for customers of the incumbent LEC and customers of competing LEC.

26. Viewed in this context, not only ought the mean be equal across customer groups, therefore, but the variance as well. In fact, for variables which are not normally distributed, higher-order moments should also be equal across customer groups (e.g., skew, kurtosis).

27. For variables which are normally distributed, therefore, it would be reasonable to test the equality of both the mean and the variance for indication of possible discrimination. For variables which are not normally distributed, non-parametric tests of the equality of the distributions may be appropriate. These are described below.

2. Non-Parametric Tests

28. If the shape of the distribution of the data is not known, or cannot be reasonably assumed, then it would be necessary to use non-parametric techniques to test the significance of differences across two samples. The exception to this rule is the test for differences in means, as described above, because of the central limit therein.

29. Non-parametric, or distribution-free, techniques do not rely on the distribution of the population, but are instead based on order statistics and can be applied to continuous and discrete random variables. Rather than directly testing the simple difference between the means of two samples, non-parametric tests will determine whether the population locations are different (i.e., whether the two samples were drawn from different populations).

30. These tests are generally simpler and less time-consuming than the parametric tests described above. In addition, because no assumptions are required with regard to the underlying population distribution(s), non-parametric tests can typically be run using less data than the parametric tests described above.

31. Parametric tests are more powerful, however, in the sense that they provide for a greater level of precision in calculating significance. In addition, when dealing with data that comes from a normal distribution, parametric tests are more accurate than non-parametric tests. For this reason, I recommend using the parametric tests in the first stage analysis. In addition, where sample sizes are large enough and distributional assumptions are met, parametic tests should be used in second stage analyses as well.

3. Bootstrapping

32. Another technique that is sometimes used when sample size is limited and re-estimation of the sample is not possible (i.e., it is not possible to simply go out and obtain a new sample, as would be the case here) is bootstrapping. Bootstrapping is a data-based simulation method for statistical inference, which involves repeated sampling from the sample.

33. As a simple example, suppose there is a sample of 10 observations. Using the bootstrap method, one will draw a sample of 10 observations with replacement (i.e., a data point may be selected more than once) from the original sample. This process would then be repeated a large number of times, perhaps 1000 times. The sample mean would then be calculated for each of these 1000 samples. These 1000 replicates would then be used to

make inferences. For example, an approximate 95% confidence interval could be derived by taking the 25th and 975th largest of the 1000 replicates.

34. Bootstrap methods are intended to simplify complicated calculations. They are, however, extremely computer-intensive techniques. While this method may be useful in second stage analyses, I recommend the use of more common statistical techniques for the first-stage parity analysis.

4. Extreme Value Theory

35. Extreme value theory is a relatively new area of statistics. Rather than evaluating the equality of the means of two distributions, this field concerns itself with detecting differences in extreme values. The theory here is to consider what the likely extreme values are for a particular variable and to compare these extreme values over different samples.

36. For example, suppose that the customers of an incumbent LEC have to wait an average of 2.6 hours for phone service to be restored, while the customers of a competing LEC have to wait an average of 2.7 hours for phone service to be restored. This difference is not likely to be significant statistically and even if it is significant, it is not clear that this difference is material enough to warrant further investigation.

37. Another measure worth considering, however, is the maximum length of time a customer has to wait. Suppose, for example, that customers of the incumbent LEC never have to wait for more than 4 hours, while customers of the competing LEC may have to wait as long as 24 hours. Clearly, this will affect the ability of the competing LEC to acquire and retain customers. If the number of customers affected at the extreme is relatively small,

however, such possible discrimination may not show up in a simple test of the means. It would be more likely to show up in a test of variances, as suggested by the Commission for this very reason. The field of extreme-value theory may provide another way to identify possible situations of this nature. However, for the first stage analysis, I recommend the use of the tests of variances to identify these occurrences. The tests of variances as described above are well-known and widely accepted.

38. Each time discrimination is to be tested, all possible tests should be conducted. The number of variables exhibiting "extreme" values can then be used to determine overall parity of performance.

5. Continuous Variables and the Use of 95 Percent Confidence Intervals.

39. Ameritech agrees with the general framework presented by AT&T for determining possible disparity in the first stage of the analysis. There is a random component to the performance of the incumbent LEC that is observed on any measure. The statistical tests to determine parity must recognize this random aspect. As AT&T acknowledges, statistical tests based on a 95 percent confidence interval will falsely indicate a disparity in service in 5 percent of the cases where parity actually exists. Put another way, in a large number of tests across multiple performance measures, 5 percent of the measures may appear to show significant departures from parity on a statistical basis even when parity actually exists.

40. This means that in measurement periods in which parity exists on all performance measures we would expect to observe up to five percent of performance measures exhibiting lack of parity. This raises the question of how many disparate performance

measures should be observed before it is determined that there is a potential lack of parity in a given period. Certainly the observation of five disparate measures is too few: that will result in a finding of disparity roughly half of the time even when there is actual parity on all measures.

41. AT&T has suggested that in general the 95 percent confidence interval is appropriate. In this case the 95 percent confidence interval means that the test criteria should be established so that overall parity would be rejected 5 percent of the time even when parity actually exists. For example, assume that there are 99 tests for parity in a given period (three for each of 33 performance measures). Across these 99 tests, statistical theory shows that in 5 percent of the test periods one would expect to observe more than 8 measures exhibiting an apparent lack of parity even when there is complete parity.

42. Obviously the number of acceptable disparate tests will depend on the number of performance measures tested. The exact number can be determined using the binomial distribution based on the number of parity tests performed in each test period. For example if 100 parity tests are performed, more than 9 percent of the tests would need to exhibit lack of parity before one could be 95 percent confident that the observed disparity may be due to more than random chance. Alternatively, if 50 parity tests are performed, more than 10 percent of the tests would need to exhibit lack of parity. In both examples, the general approach remains the same: The number of performance measures allowed to exhibit lack of parity in any given test period will be set to establish a 95 percent confidence interval.

43. It is important to recognize that many of the tests for parity suggested in the Notice are highly correlated to one another. This correlation results from two potential sources. First, some of the measures are based on the same underlying performance function by the incumbent LEC. For example, Average Completion Interval will be highly correlated with Percentage of Due Dates Missed. If disparity is observed on one of these measures, it is very likely that it will be observed on the other measure. Second, since a single performance measure will often be disaggregated into multiple parity tests, it is likely that lack of parity on one disaggregated unit will be associated with lack of parity in other disaggregations.

44. For both of these reasons, random variations in the observed performance of the incumbent LEC may lead to multiple measures exhibiting an apparent lack of parity, resulting in a false finding of overall disparity. Such potential false alarms are acceptable as long as the parity tests are used as a device to trigger further investigation and analysis and to determine where potential adjustments or corrections should be made. However, if a first-level finding of an indication of possible disparity were an automatic trigger for enforcement action, this potential for correlation among parity tests and across performance measures directed at a single function of the incumbent LEC would have to be eliminated. This would require eliminating redundant or closely related performance measures from those suggested by the Commission.

45. This set of tests would establish a threshold standard, described by the Commission on page 47 of the Notice (paragraph 121). A failure to exceed this threshold standard would result in a definitive ruling of non-discrimination.

46. Exceeding this threshold standard should not, however, result in an automatic ruling of discrimination. Instead, an incumbent LEC which exceeds this threshold standard would be subject to investigation and more intensive analysis as described in "Step 2" below.

B. Step 2: Determination of Discrimination

47. If possible discrimination is indicated in Step 1 above, the variables for which discrimination is suspected should be further investigated in order to attempt to determine as definitively as possible whether or not discrimination exists.

Put another way, the basic technique addressed by the Notice answers the following question:

Is there a statistically significant difference between m1 and m2?

The more pertinent question, however, is: Is there evidence of discrimination? In other words, is there a statistically significant difference between m1 and m2 after we control for the effects of factors which are not discrimination, but which may cause differences between m1 and m2?

48. The variable for which discrimination may be indicated should be fully specified as a function of all variables which may possibly influence it. This should include a measure of whether the relevant customer was a customer of the incumbent LEC. If the fact that the customer was a customer of the incumbent LEC is found to be a statistically significant determinant of the variable at issue then discrimination can be asserted with a relatively high degree of certainty.

49. Possible techniques available for this stage are discussed below.